Efficiently Learning the Graph for Semi-supervised Learning

Dravyansh Sharma[1]; Maxwell Jones[1]
[1]Carnegie Mellon University

Introduction

We consider graph-based approaches for semi-supervised learning on large datasets. Known techniques to improve efficiency typically involve an approximation of the graph regularization objective, however:

- The graph is assumed to be known or constructed with heuristic hyperparameter values
- They do not provide a principled approximation guarantee for learning over the full unlabeled dataset.

We propose algorithms that overcome both limitations.

- We learn the graph $G(\sigma)$ via algorithms that can exhaustively search a continuous parameter space
- We give an online learning framework to learn the graph efficiently online
- We speedup our algorithm by using the Conjugate Gradient method as an approximate matrix inversion method
- We give guarantees of our algorithm given approximate feedback
- We observe significant (~10-100x) speedup over prior work

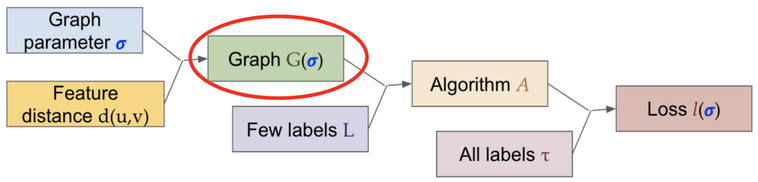


Figure 1. A visual of the semi-supervised learning setup. We learn graph $G(\sigma)$ by finding optimal σ

Online Learning Setup

Dispersion. Consider a parameter space $\mathcal{C}$, a sequence of random loss functions $l_1, \dots, l_T$: $\mathcal{C} \to [0,1]$, and a constant β. We consider this random sequence β-dispersed for Lipschitz constant L if, for all $\epsilon \geq T^{-\beta}$, each pair of points at distance ϵ in $\mathcal{C}$ is not L-Lipschitz for at most $\tilde{O}(\epsilon \mathrm{T})$ functions in expectation.

An online optimization problem with loss functions $l_1, l_2, \dots$ is said to have (ϵ, γ)**-approximate semi-bandit feedback** with system size M if there is a partition $\tilde{A}_t^{(1)}, \dots \tilde{A}_t^{(m)}$ of the parameter space $\mathcal{P} \subset \mathbb{R}^d$, such that if the learner plays point $\rho_t \in \tilde{A}_t^{(i)}$, she observes approximate feedback set $\tilde{A}_t^{(i)}$ and approximate loss $\tilde{l}_t(\rho)$, which is within γ of the loss for all $\rho_t \in \tilde{A}_t^{(i)}$ except for some subset of $\tilde{A}_t^{(i)}$ with volume at most ϵ.

Algorithm 1: Approximate Continuous exp3-set (λ)

Input: step size $\lambda \in [0, 1]$.
Initialize $w_1(\rho) = 1$ for all $\rho \in \mathcal{P}$.
for $t = 1, \dots, T$ **do**
  Sample ρ_t according to $p_t(\rho) = \frac{w_t(\rho)}{\int_{\mathcal{P}} w_t(\rho) d\rho}$.
  Play ρ_t and suffer loss $l_t(\rho_t)$.
  Observe (γ, ϵ)-approximate feedback $\tilde{l}_t(\rho)$ over set $\tilde{A}_t$ with $\rho_t \in \tilde{A}_t$
  Update $w_{t+1}(\rho) = w_t(\rho) \exp(-\lambda \hat{l}_t(\rho))$, where $\hat{l}_t(\rho) = \frac{\mathbf{I}\{\rho \in \tilde{A}_t\}}{\int_{\tilde{A}_t} p_t(\rho) d\rho} \tilde{l}_t(\rho)$.

Online Learning Guarantees

Regret bound for Algorithm 1. Suppose $l_1, \dots, l_T$: $\mathcal{C} \to [0,1]$ is a sequence of loss functions that is β-dispersed, and the domain $\mathcal{C} \subset \mathbb{R}^d$ is contained in a ball of radius R. Algorithm 1 achieves expected regret

$$\tilde{O}\left(\sqrt{dMT \log RT} + T^{1-\min\{\beta,\beta'\}}\right)$$

with access to (ϵ, γ)-approximate semi-bandit feedback with system size M, provided $\gamma \leq \text{volume}\left(\mathfrak{B}\left(T^{-\beta}\right)\right) T^{-\beta}$, where $\mathfrak{B}(r)$ is a d-ball of radius r.

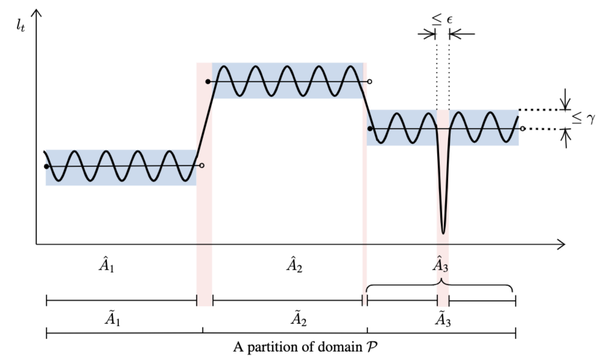


Figure 2. A depiction of a (ϵ, γ)**-approximate semi-bandit feedback** with system size 3.

Graph Learning Setup

Gaussian nearest neighbors. Consider a data space $\mathcal{X}$ and distance function $d: \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$. Further, let N'_k denote a subset of $\mathcal{X} \times \mathcal{X}$, where $(u, v) \in N'_k$ indicates u is a k-nearest neighbor to v AND v is a k nearest neighbor to u under metric $d(\cdot,\cdot)$. Finally, construct graph $G(k, \sigma)$ with edge weights:

$$w(u, v) = e^{\frac{d(u,v)^2}{\sigma^2}} \mathbb{I}[(u, v) \in N'_k]$$

for all instances $u, v \in \mathcal{X}$. We let $\mathcal{H}_{k,\sigma}$ denote a set of functions that take in some $k \in [K], \sigma \in \mathbb{R}$ and output loss w.r.t. label predictions on graph $G(k, \sigma)$.

Pseudo-dimension of $\mathcal{H}_{k,\sigma}$. We show that the pseudo-dimension of $\mathcal{H}_{k,\sigma}$ is $O(K + \log n)$, where n denotes the number of nodes in the graph, and the labeling algorithm is the mincut approach of Blum and Chawla [2001].

Algorithm 2: Approximate Feedback Set (ϵ, η)

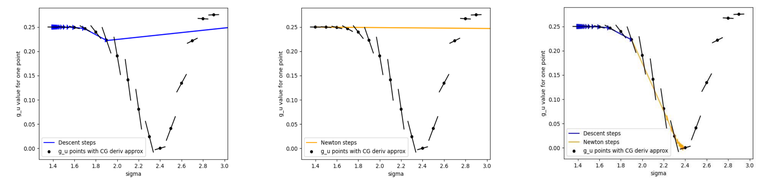
Input: Graph G with unlabeled nodes U, labels f_L, query parameter σ_0, error tolerance ϵ, learning rate η, algorithm $\mathcal{A}$ to estimate soft labels and derivatives at any σ
Output: Estimates for piecewise constant interval containing σ_0, and function value at σ.
for all $u \in U$ **do**
  while $|\sigma_{n+1} - \sigma_n| \geq \epsilon$ **do**
    Compute $f_{u,\epsilon}(\sigma), \frac{\partial f_u}{\partial \sigma}_{\epsilon}$ as $\mathcal{A}(G, f_L, u, \sigma_n, \epsilon)$, where f is the soft label function
    Set $g_u(\sigma_n) = (f_{u,\epsilon}(\sigma_n) - \frac{1}{2})^2$
    Compute Gradient Descent and Newton's method steps $\xi_{\text{GD}}, \xi_{\text{Newton}}$ for $g_u(\sigma_n)$
    Step σ_{n+1} towards σ^*, where $g_u(\sigma^*) = 0$ by setting $\sigma_{n+1} = \sigma_n - \min\{\xi_{\text{GD}}, \xi_{\text{Newton}}\}$.
  $\sigma_l = \min\{\sigma_l, \sigma_{n+1}\}, \sigma_h = \max\{\sigma_h, \sigma_{n+1}\}, n \leftarrow n + 1$.
return $[\sigma_l, \sigma_h]$, $f_\epsilon(\sigma_0)$.

Graph Learning Guarantees

Complexity bound for Algorithm 2. Given an algorithm for computing ϵ-approximate soft labels and gradients for the efficient semi-supervised learning algorithm of Delalleau et al. [2005], Algorithm 2 computes (ϵ, ϵ)-approximate semi-bandit feedback for loss $l(\sigma)$ in time

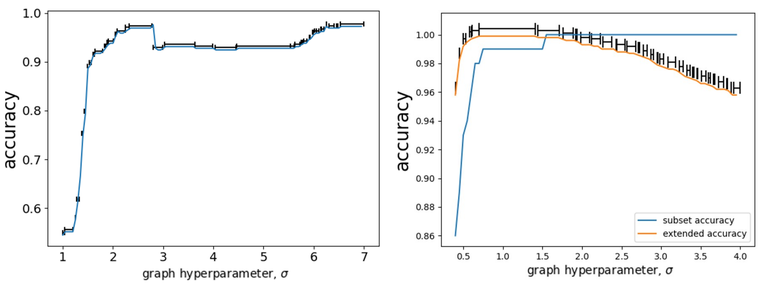
$$O\left(\left|E_{G_{\widetilde{U}}}\right| n\sqrt{\kappa(\mathcal{L}_A)} \log\left(\frac{\lambda\left((|L| + |\widetilde{U}|)\Delta\right)}{\epsilon \sigma_{min} \lambda_{min}(\mathcal{L}_A)}\right) \log\log\frac{1}{\epsilon}\right)$$

Here $E_{G_{\widetilde{U}}}$ represents edges w.r.t. graph of a small subset of unlabeled nodes $\widetilde{U}$, $\mathcal{L}_A$ represents grounded graph Laplacian of graph A used to determine labels for $\widetilde{U}$, and Δ represents the size of the interval for parameter σ. Notice that this bound is linear w.r.t. n, assuming a well conditioned graph.



(a) Gradient Descent Only (b) Newton's Method Only (c) Our Method

Figure 3. An instance where finding local minima of $g_u(\sigma) = \left(f_u(\sigma) - \frac{1}{2}\right)^2$ is challenging. Our method (taking the min of Gradient Descent/Newton's method steps) finds the minima.



(a) Subset of FashionMNIST k=6, 30 |L|=30, |U|=300

(b) Subset of USPS k=6, |L|=10, $|\widetilde{U}|$=100, |U|=1000

Figure 4. Loss intervals calculated with approximate. Black intervals are estimated constant loss.

Results/Discussion

- We show a formal separation in the learning-theoretic complexity of sparse and dense graph families.
- We show how to approximately learn the best graphs from the sparse families efficiently using the conjugate gradient method.
- We provide an online learning framework that can be used to learn the graph efficiently online with sub-linear regret, under mild smoothness assumptions
- We implement our approach and demonstrate significant (~10-100x) speedups over prior work.

References

1. Avrim Blum and Shuchi Chawla. Learning from labeled and unlabeled data using graph mincuts. In International Conference on Machine Learning (ICML), 2001.
2. Olivier Delalleau, Yoshua Bengio, and Nicolas Le Roux. Efficient non-parametric function induction in semi-supervised learning. In International Workshop on Artificial Intelligence and Statistics (AISTATS), pages 96–103. PMLR, 2005.
3. Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In International Conference on Machine Learning (ICML), pages 912–919, 2003.
4. Maria-Florina Balcan and Dravyansh Sharma. Data driven semi-supervised learning. Advances in Neural Information Processing Systems (NeurIPS), 34, 2021.